AD-A093 634 WISCONSIN UNIV-MADISON MATHEMATICS RESEARCH CENTER F/G 12/1
SMOOTHING ESTIMATION OF STOCHASTIC PROCESSES. PART II. TWO FILT--ETC(U)
SEP 80 V SOLO DAAG29-80-C-0041
UNCLASSIFIED MRC-TSR-2120 NL

1 OF 1

END
DATE FILMED
2-84
DTIC

MRC Technical Summary Report #2120

SMOOTHING ESTIMATION OF STOCHASTIC PROCESSES - PART II: TWO FILTER FORMULAe

V. Solo

AD A093634

(1)

LEVEL

Mathematics Research Center
University of Wisconsin—Madison
610 Walnut Street
Madison, Wisconsin 53706

September 1980

Received May 23, 1980

DDC FILE COPY

Jee 1473

JAN 1981

Approved for public release
Distribution unlimited

Sponsored by

U. S. Army Research Office
P. O. Box 12211
Research Triangle Park
North Carolina 27709

80 12 22 045

1

UNIVERSITY OF WISCONSIN - MADISON
MATHEMATICS RESEARCH CENTER

SMOOTHING ESTIMATION OF STOCHASTIC PROCESSES
PART II: TWO FILTER FORMULAE

V. Solo

Technical Summary Report #2120
September 1980

ABSTRACT

Some simple derivations of two-filter-like formulae (in the smoothing problem of linear estimation) are given for general nonstationary process. It then becomes clear how a wide sense Markovian assumption is required to give the formulae a backwards filter interpretation.

AMS(MOS) Subject Classification: 93E10, 60G35

Key words: Linear estimation, smoothing, filtering,
Markov process, covariance kernel, innovations

Work Unit Number 4 - Statistics and Probability

Sponsored by the United States Army under Contract No. DAAG29-80-C-0041.

SIGNIFICANCE AND EXPLANATION

The estimation of one process from measurements on another related process is a problem that arises in many areas such as Time Series Analysis, Econometrics, Communications Engineering and Control Engineering. Particularly in the Engineering applications there is a great interest in various computational forms of the algorithms proposed to solve the above problem. The aim in this article is to give simple derivations of some of these algorithms thus revealing how they apply to general nonstationary processes. This facilitates an understanding of what minimal assumptions are needed for their full utility.

The responsibility for the wording and views expressed in this descriptive summary lies with MRC, and not with the author of this report.

SMOOTHING ESTIMATION OF STOCHASTIC PROCESSES
PART II: TWO FILTER FORMULA

V. Solo

1. INTRODUCTION. Recently a number of authors have discussed various types of smoothing formulae in varying levels of generality (so far as the signal and observed process models are concerned): see Kailath and Frost [6], Ljung and Kailath [9], Lainiotis [8]. An ongoing problem has been the understanding of the two-filter formulae: Mayne [10], Fraser [3], Mehra, Badawi et. al. [1]. In this article these two-filter results and some new ones are derived in a simple way in a very general setting (for arbitrary nonstationary processes). It turns out however that only if a wide-sense (i.e. second order) Markovian assumption is added can one of the filters be viewed as a backwards filter. The remainder of the paper is organized as follows. Section 2 recalls some smoothing formulae that apply to both continuous and discrete observations. Section 3 discusses two types of two-filter-like formulae for general nonstationary processes. In Section 4 one of the filters is shown to be a backwards least squares estimate provided a wide sense Markovian assumption is satisfied. Section 5 contains a derivation of some backwards filters. In Section 6 some additional two-filter-like formulae are given. The final section is a conclusion.

2. PRELIMINARIES. Consider the linear estimation of an n_x dimensional process $\underline{x}(t)$ from measurements on a related n_y dimensional process $\underline{y}(t)$. In the first instance suppose $\underline{y}(t)$ is measured in discrete time at points $0 < t_1 < t_2 < \ldots < t_N < T$ over an interval $[0,T]$; collect these observations into a vector $\underline{\mathcal{Y}}_T$ and assume the covariance matrix $E(\underline{\mathcal{Y}}_T \underline{\mathcal{Y}}_T')$ is positive definite. Now for any t in $[0,T]$ $\underline{\mathcal{Y}}_T$ is comprised of two vectors $\underline{\mathcal{Y}}_t$, $\underline{\mathcal{Y}}_{tT}$ consisting of the data over the intervals $[0,t]$ and $(t,T]$ respectively. Let us denote the linear least squares predictor by

$$\hat{\underline{x}}(t|T) \quad \text{or} \quad \hat{E}(\underline{x}(t)|\underline{\mathcal{Y}}_T)$$

where $\hat{E}$ denotes wide sense conditional expectations or projection. (see Parzen [12, p309]; Doob [2, p150]). Now $\hat{\underline{x}}(t|T)$ is defined by the orthogonality condition

Sponsored by the United States Army under Contract No. DAAG29-80-C-0041.

$$E(\underline{x}(t) - \underline{\hat{x}}(t,T))\mathcal{Y}'_T = \underline{0}\ .$$

Thus

$$\underline{\hat{x}}(t,T) = \hat{E}(\underline{x}(t)|\mathcal{Y}_T) = E(\underline{x}(t)\mathcal{Y}'_T)E(\mathcal{Y}_T\,\mathcal{Y}'_T)^{-1}\mathcal{Y}_T\ .$$

Also denote $\hat{\mathcal{Y}}_{tT} = \hat{E}(\mathcal{Y}_{tT}|\mathcal{Y}_t)$ and $\tilde{\mathcal{Y}}_{tT} = \mathcal{Y}_{tT} - \hat{\mathcal{Y}}_{tT}$. Then observe

$$\begin{aligned}\underline{\hat{x}}(t|T) &= \hat{E}(\underline{x}(t)|\mathcal{Y}_T)\\ &= \hat{E}(\underline{x}(t)|\mathcal{Y}_t,\ \mathcal{Y}_{tT})\\ &= \hat{E}(\underline{x}(t)|\mathcal{Y}_t,\ \tilde{\mathcal{Y}}_{tT})\\ &= \hat{E}(\underline{x}(t)|\mathcal{Y}_t) + \hat{E}(\underline{x}(t)|\tilde{\mathcal{Y}}_{tT}) && \text{by orthogonality}\\ &= \underline{\hat{x}}(t|t) + \underline{\gamma}(t|T) && \text{say.}\end{aligned}$$

Now define

$$\underline{P}(t|T) = E(\underline{x}(t) - \underline{\hat{x}}(t|T))(\underline{x}(t) - \underline{\hat{x}}(t|T))'$$

and denote $\underline{P}(t) = \underline{P}(t|t)$. Then call

$$\underline{\lambda}(t|T) = \underline{P}^{-1}(t)\underline{\gamma}(t|T).$$

Thus we can write

$$\hat{x}(t|T) = \underline{\hat{x}}(t|t) + \underline{P}(t)\underline{\lambda}(t|T) \qquad (1)$$

Also observe that

$$E[\underline{\hat{x}}(t|t)\underline{\lambda}'(t|T)] = \underline{0} \qquad (2)$$

So we can then find

$$\begin{aligned}\underline{P}(t) &= E(\underline{x}(t) - \underline{\hat{x}}(t|t))(\underline{x}(t) - \underline{\hat{x}}(t|t)'\\ &= \underline{P}(t|T) + \underline{P}(t)\mathcal{Q}(t|T)\underline{P}(t)\end{aligned}$$

or

$$\underline{P}(t|T) = \underline{P}(t) - \underline{P}(t)\mathcal{Q}(t|T)\underline{P}(t) \qquad (3)$$

where we have introduced

$$\mathcal{Q}(t|T) = E[\underline{\lambda}(t|T)\underline{\lambda}'(t|T)] \qquad (4)$$

Equations (1)-(4) describe general continuous-discrete estimation formulae.

Now suppose $\underline{y}(t)$ is measured continuously over the interval $[0,T]$. Assume $\underline{y}(t)$ has finite variance and a positive definite covariance Kernel. (Thus in a signal plus noise model we are writing $d\underline{y}(t) = \underline{S}(t)dt + d\underline{W}(t)$.)

Now use the symbol $\mathcal{Y}_T$ to denote the Hilbert space spanned by $\underline{y}(\sigma)$, σ in $[0,T]$ with inner product $\langle u,v \rangle = E(uv)$ (for random variables $u,v \in \mathcal{Y}_T$). That is $\mathcal{Y}_T$ consists of all random variables that are finite linear combinations (of the form $\sum_i \cdot \underline{a}_i' \underline{y}(t_i)$, t_i in $[0,T]$) of $\underline{y}(\sigma)$ for σ in $[0,T]$ or limits in mean square of such linear combinations. Also call $\mathcal{Y}_t$, $\mathcal{Y}_{tT}$ the Hilbert subspaces spanned by $\underline{y}(\sigma)$, σ in $[0,t)$, $[t,T]$ respectively. The linear least squares estimate $\hat{\underline{x}}(t|T)$ is the vector whose ith component is the unique projection $\hat{E}(x_i(t)|\mathcal{Y}_T)$ that satisfies

$$E(x_i(t) - \hat{E}(x_i(t)|\mathcal{Y}_T))\underline{y}'(\sigma) = 0 , \quad \text{for all } \sigma \text{ in } [0,T]$$

with a slight abuse of notation denote $\hat{\underline{x}}(t|T)$ by $\hat{E}(\underline{x}(t)|\mathcal{Y}_T)$. Denote by $\hat{\mathcal{Y}}_{tT}$(= "$\hat{E}(\mathcal{Y}_{tT}|\mathcal{Y}_t)$") to be the Hilbert subspace of $\mathcal{Y}_t$ spanned by $\hat{E}(\underline{y}(s)|\mathcal{Y}_t)$ s in $[t,T]$. Then $\tilde{\mathcal{Y}}_{tT}$(="$\mathcal{Y}_{tT} - \hat{\mathcal{Y}}_{tT}$") is the orthogonal complement of $\mathcal{Y}_t$ in $\mathcal{Y}_T$ namely the Hilbert subspace of $\mathcal{Y}_T$ spanned by $\underline{y}(s) - \hat{E}(\underline{y}(s)|\mathcal{Y}_t)$, s in $[t,T]$.

Then observe exactly as before

$$\begin{aligned}
\hat{\underline{x}}(t|T) &= \hat{E}(\underline{x}(t)|\mathcal{Y}_T) \\
&= \hat{E}(\underline{x}(t)|\mathcal{Y}_t,\mathcal{Y}_{tT}) \\
&= \hat{E}(\underline{x}(t)|\mathcal{Y}_t,\tilde{\mathcal{Y}}_{tT}) \\
&= \hat{E}(\underline{x}(t)|\mathcal{Y}_t) + \hat{E}(\underline{x}(t)|\tilde{\mathcal{Y}}_{tT}) \\
&= \hat{\underline{x}}(t|t) + \underline{\gamma}(t|T).
\end{aligned}$$

With the same definitions as before it easily follows that (1)-(4) hold also in continuous time. These relations have been previously given by Kailath and Frost [6] for continuous time processes possessing an innovations process. See also Kailath and Geesey [7].

Since much of the ensuing argument depends only on (1)-(4) the discrete, continuous-discrete, and continuous cases can be given a joint treatment.

3. <u>TWO-FILTER FORMULAE</u>. (A) <u>The basic idea</u>. The idea of a two-filter formula is to compute the smoothed estimate at time t as a sum of two estimates; one using data in $[0,t)$ the other using data in $[t,T)$. If these estimates are least squares they ought to be orthogonal in some sense. Thus the formula could be of the form

$$\hat{\underline{x}}(t|T) = \underline{P}(t|T)(\underline{P}^{-1}(t)\hat{\underline{x}}(t|t) + \underline{P}_B^{-1}(t|T)\hat{\underline{x}}_B(t|T))$$

where the subscript B denotes backward and

$$\underline{P}_B(t|T) = E[\underline{x}(t) - \hat{\underline{x}}_B(t|T)][\underline{x}'(t) - \hat{\underline{x}}_B(t|T)].$$

Of course as pointed out in (2) two terms in the basic decomposition (1) have an orthogonality property but just what the "backwards" orthogonality should be in general is not clear; there are indeed two possibilities

$$E(\underline{x}(t) - \hat{\underline{x}}_B(t|T))\underline{x}'(t) = \underline{0}$$

or

$$E(\underline{x}(t) - \hat{\underline{x}}_B(t|T))\hat{\underline{x}}_B'(t|T) = \underline{0}\ .$$

It turns out that both these lead to satisfactory expressions. *The required two-filter-like formulae will be obtained by reorganizing the basic formula (1).*

(B) <u>The first two-filter form</u>. First apply the Matrix Inversion Lemma to (4) to see

$$\underline{P}^{-1}(t|T) = \underline{P}^{-1}(t) + (\underline{\mathcal{Q}}^{-1}(t|T) - \underline{P}(t))^{-1}. \tag{4a}$$

So define $\underline{P}_\rho(t|T)$ (ρ stands for reverse) by

$$\underline{\mathcal{Q}}^{-1}(t|T) = \underline{P}_\rho(t|T) + \underline{P}(t) \tag{5}$$

(notice that as $t \to T$, $\underline{\mathcal{Q}}(t|T) \to \underline{0}$ so $\underline{P}_\rho(t|T) \to \infty$) so that (4a) is rewritten

$$\underline{P}^{-1}(t|T) = \underline{P}^{-1}(t) + \underline{P}_\rho^{-1}(t|T). \tag{6}$$

Now multiply this through (1) to find

$$\underline{P}^{-1}(t|T)\hat{\underline{x}}(t|T) = \underline{P}^{-1}(t)\hat{\underline{x}}(t|t) + \underline{P}_\rho^{-1}(t|T)(\hat{\underline{x}}(t|t) +$$
$$\underline{P}_\rho(t|T)(\underline{P}^{-1}(t) + \underline{P}_\rho^{-1}(t|T))\underline{P}(t)\underline{\lambda}(t|T)))$$
$$= \underline{P}^{-1}(t)\hat{\underline{x}}(t|t) + \underline{P}_\rho^{-1}(t|T)(\hat{\underline{x}}(t|t) + (\underline{P}_\rho(t|T) + \underline{P}(t))\underline{\lambda}(t|T))$$
$$= \underline{P}^{-1}(t)\hat{\underline{x}}(t|t) + \underline{P}_\rho^{-1}(t|T)(\hat{\underline{x}}(t|t) + \underline{\mathcal{Q}}^{-1}(t|T)\underline{\lambda}(t|T))$$

by (5)

We are thus led to introduce

$$\hat{\underline{x}}_{\rho}(t|T) = \hat{\underline{x}}(t|t) + \underline{\mathcal{O}}^{-1}(t|T)\underline{\lambda}(t|T) \tag{7a}$$

so we can write

$$\hat{\underline{x}}(t|T) = \underline{P}(t|T)(\underline{P}^{-1}(t)\hat{\underline{x}}(t|t) + \underline{P}_{c}^{-1}(t|T)\hat{\underline{x}}_{\rho}(t|T)) \tag{7b}$$

and calculate

$$\begin{aligned} &E(\underline{x}(t) - \hat{\underline{x}}_{\rho}(t|T))(\underline{x}'(t) - \hat{\underline{x}}'_{\rho}(t|T)) \\ &\quad = E(\underline{x}(t) - \hat{\underline{x}}(t|t) - \underline{\mathcal{O}}^{-1}(t|T)\underline{\lambda}(t|T))(\underline{x}'(t) - \hat{\underline{x}}'(t|t) - \underline{\mathcal{O}}^{-1}(t|T)\underline{\lambda}(t|T))' \\ &\quad = \underline{P}(t) - E(\underline{x}(t)\underline{\lambda}'(t|T))\underline{\mathcal{O}}^{-1}(t|T) - \underline{\mathcal{O}}^{-1}(t|T)E(\underline{\lambda}(t|T)\underline{x}'(t)) + \underline{\mathcal{O}}^{-1}(t|T) \end{aligned}$$

in view of (2) and the definition of $\underline{P}(t)$.

Next observe

$$\begin{aligned} E(\underline{x}(t)\underline{\lambda}'(t|T)) &= E[\underline{x}(t)(\hat{\underline{x}}(t|T) - \hat{x}(t))']\underline{P}^{-1}(t) \\ &= -(\underline{P}(t|T) - \underline{P}(t))\underline{P}^{-1}(t) \\ &= +\underline{P}(t)\underline{\mathcal{O}}(t|T) \qquad \text{by (3)} \end{aligned} \tag{8}$$

Thus the above expression becomes

$$\begin{aligned} &E(\underline{x}(t) - \hat{\underline{x}}_{\rho}(t|T)(\underline{x}(t) - \hat{\underline{x}}_{\rho}(t|T))' \\ &\quad = \underline{P}(t) - \underline{P}(t) - \underline{P}(t) + \underline{\mathcal{O}}^{-1}(t|T) \\ &\quad = \underline{P}_{\rho}(t|T) \qquad \text{by (5)}\ . \end{aligned} \tag{9}$$

Also observe that $\hat{\underline{x}}_{\rho}(t|T)$ satisfies the following orthogonality properties

$$E(\underline{x}(t) - \hat{\underline{x}}_{\rho}(t|T))\underline{x}'(t) = \underline{0} \tag{10a}$$

$$E(\underline{x}(t) - \hat{\underline{x}}_{\rho}(t|T))\hat{\underline{x}}'(t|t) = \underline{0}\ . \tag{10b}$$

The first follows since

$$\begin{aligned} E(\underline{x}(t) - \hat{\underline{x}}_{\rho}(t|T))\underline{x}'(t) &= E[\underline{x}(t) - \hat{\underline{x}}(t|t) - \underline{\mathcal{O}}^{-1}(t|T)\underline{\lambda}(t|T)]\underline{x}'(t) \\ &= \underline{P}(t) - \underline{\mathcal{O}}^{-1}(t|T)\underline{\mathcal{O}}(t|T)\underline{P}(t) \qquad \text{by (8)} \\ &= \underline{0} \end{aligned}$$

while

$$\begin{aligned} E(\underline{x}(t) - \hat{\underline{x}}_{\rho}(t|T))\hat{\underline{x}}'(t|t) &= E(\underline{x}(t) - \hat{\underline{x}}(t|t) - \underline{\mathcal{O}}^{-1}(t|T)\underline{\lambda}(t|T))\hat{\underline{x}}'(t|t) \\ &= \underline{0} - \underline{0} \\ &= \underline{0} \end{aligned}$$

by (2) and the orthogonality property of $\hat{\underline{x}}(t|t)$.

Equations (7a), (7b), (9), (5), (6), (10) describe pseudo two-filter formulae for the smoothing problem. (Note that these results have also been obtained recently by Badawi et. al. [1] for a State Space Model by very elaborate argument.) The descriptor "pseudo" refers to the fact that it has not been demonstrated whether $\hat{\underline{x}}_\rho(t|T)$ can be computed by a backwards filter. This point is discussed in Section 4 below. First however we investigate the estimate, call it $\hat{\underline{x}}_\beta(t|T)$, that satisfies the other type of backwards orthogonality

$$E(\underline{x}(t) - \hat{\underline{x}}_\beta(t|T))\hat{\underline{x}}'_\beta(t|T) = \underline{0} . \tag{11}$$

(C) <u>The second two-filter form.</u> Let us denote $\underline{\pi}(t) = E(\underline{x}(t)\underline{x}'(t))$ and look for $\hat{\underline{x}}_\beta(t|T)$ in the form

$$\hat{\underline{x}}_\beta(t|T) = \underline{M}(t)\ \hat{\underline{x}}_\rho(t|T)$$

where $\underline{M}(t)$ is to be chosen to ensure (11) holds.

Consider then

$$\begin{aligned}\underline{0} &= E(\underline{x}(t) - \hat{\underline{x}}_\beta(t|T))\hat{\underline{x}}'_\beta(t|T)\\ &= E(\underline{x}(t) - \underline{M}(t)\hat{\underline{x}}_\rho(t|T))\hat{\underline{x}}'_\rho(t|T)\underline{M}'(t)\end{aligned}$$

implying

$$\underline{M}(t) = E[\underline{x}(t)\hat{\underline{x}}'_\rho(t|T)]E(\hat{\underline{x}}_\rho(t|T)\hat{\underline{x}}'_\rho(t|T))^{-1}$$

so that we have the interesting interpretation

$$\hat{\underline{x}}_\beta(t|T) = \hat{E}(\underline{x}(t)|\hat{\underline{x}}_\rho(t|T)) .$$

To find a more informative expression for $\underline{M}(t)$ continue with (10a), (10b) which imply

$$E(\underline{x}(t)\hat{\underline{x}}'_\rho(t|T)) = E(\underline{x}(t)\underline{x}'(t)) = \underline{\pi}(t) \tag{12a}$$

$$-\underline{P}_\rho(t|T) = E(\underline{x}(t) - \hat{\underline{x}}_\rho(t|T))\hat{\underline{x}}'_\rho(t|T)$$

so that

$$E(\hat{\underline{x}}_\rho(t|T)\hat{\underline{x}}'_\rho(t|T)) = \underline{\pi}(t) + \underline{P}_\rho(t|T) . \tag{12b}$$

Thus
$$\hat{\underline{x}}_\beta(t|T) = \underline{\pi}(t)(\underline{\pi}(t) + \underline{P}_\rho(t|T))^{-1}\hat{\underline{x}}_\rho(t|T) . \tag{12c}$$

Now we can introduce

$$\begin{aligned}\underline{P}_\beta(t|T) &= E(\underline{x}(t) - \hat{\underline{x}}_\rho(t|T))(\underline{x}'(t) - \hat{\underline{x}}'_\beta(t|T))\\ &= E(\underline{x}(t) - \hat{\underline{x}}_\beta(t|T))\underline{x}'(t)\\ &= \underline{\pi}(t) - \underline{\pi}(t)(\underline{\pi}(t) + \underline{P}_\rho(t|T))^{-1}\underline{\pi}(t) .\end{aligned}$$

Thus invoking the Matrix Inversion Lemma

$$\underline{P}_{\beta}^{-1}(t|T) = \underline{\pi}^{-1}(t) + \underline{P}_{\rho}^{-1}(t|T) \; . \qquad (13)$$

Observe that, from (5), as $t \to T$ $\quad \underline{P}_{\rho}(t|T) \to \infty$ while $\underline{P}_{\beta}(t|T) \to \underline{\pi}(T)$. Also from (12b)

$$E[\hat{\underline{x}}_{\beta}(t|T)\underline{x}'_{\beta}(t|T)] \to \underline{0} \qquad \text{as} \quad t \to T$$

implying that the initial condition for the computation of $\hat{\underline{x}}_{\beta}(t|T)$ is $\hat{\underline{x}}_{\beta}(T|T) = \underline{0}$. To calculate $\hat{\underline{x}}(t|T)$ using $\hat{\underline{x}}_{\beta}(t|T)$ return to (13) and find via (12c) that

$$\begin{aligned}
\underline{P}_{\rho}^{-1}(t|T)\hat{\underline{x}}_{\rho}(t|T) &= \underline{P}_{\rho}^{-1}(t|T)(\underline{\pi}(t) + \underline{P}_{\rho}(t|T))\underline{\pi}^{-1}(t)\hat{\underline{x}}_{\beta}(t|T) \\
&= (\underline{P}_{\rho}^{-1}(t|T) + \underline{\pi}^{-1}(t))\hat{\underline{x}}_{\beta}(t|T) \\
&= \underline{P}_{\beta}^{-1}(t|T)\hat{\underline{x}}_{\beta}(t|T) \qquad \text{by (13).} \qquad (14)
\end{aligned}$$

The interesting "invariance" expressed in this relation explains some of the confusion with the two-filter formulae. To summarize we collect some of the expressions together

$$\begin{aligned}
\hat{\underline{x}}(t|T) &= \hat{\underline{x}}(t|t) + \underline{P}(t)\underline{\lambda}(t|T) && (1) \\
&= \underline{P}(t|T)(\underline{P}^{-1}(t)\hat{\underline{x}}(t|t) + \underline{P}_{\rho}^{-1}(t|T)\hat{\underline{x}}_{\rho}(t|T)) && (7b) \\
&= \underline{P}(t|T)(\underline{P}^{-1}(t)\hat{\underline{x}}(t|t) + \underline{P}_{\beta}^{-1}(t|T)\hat{\underline{x}}_{\beta}(t|T)) && (15a) \\
&= \underline{P}(t|T)(\underline{P}^{-1}(t)\hat{\underline{x}}(t|t) + (\underline{P}_{\rho}^{-1}(t|T) + \underline{\pi}^{-1}(t))\hat{\underline{x}}_{\beta}(t|T)) \\
&\qquad \text{by (13).} && (15b)
\end{aligned}$$

Also

$$\underline{P}^{-1}(t|T) = \underline{P}^{-1}(t) + \underline{P}_{\rho}^{-1}(t|T). \qquad (6)$$

Expression (15b) was given by Ljing and Kailath [9, p155] for a state space model. With (11) in mind we turn now to consider when $\hat{\underline{x}}_{\beta}(t|T)$ can be computed by a backwards filter i.e. when is $\hat{\underline{x}}_{\beta}(t|T)$ a backwards estimate of $\underline{x}(t)$ based on the data in $[t,T]$.

4. BACKWARDS FILTERS AND A MARKOVIAN ASSUMPTION. Here we investigate under what conditions $\hat{x}_\beta(t|T)$ is the linear least squares estimate, $\hat{\underline{x}}_b(t|T)$, of $\underline{x}(t)$ given the data $\underline{y}$ in (t,T]. Since the least squares estimate is unique we can establish when it is also $\hat{\underline{x}}_\beta(t|T)$ by ensuring

$$E(\underline{x}(t) - \hat{\underline{x}}_\beta(t|T))\underline{y}'(s) = \underline{0} \text{ for all observation points } s \text{ in } (t,T]. \tag{16}$$

First we find a convenient expression for $\hat{\underline{x}}_\beta(t|T)$. Since $\hat{\underline{x}}_\beta(t|T)$ is to be a backwards estimate we can expect a backwards decomposition analogous to (1) say

$$\hat{\underline{x}}(t|T) = \hat{\underline{x}}_\beta(t|T) + \underline{P}_\beta(t|T)\underline{\lambda}_\beta(t|T) \tag{17a}$$

with $E(\hat{\underline{x}}_\beta(t|T)\underline{\lambda}'_\beta(t|T)) = \underline{0}$.

This is indeed possible with (See Appendix A)

$$\begin{aligned}\underline{\lambda}_\beta(t|T) &= \underline{\pi}^{-1}(t)[\hat{\underline{x}}(t|t) + (\underline{P}(t) - \underline{\pi}(t))\underline{\lambda}(t|T)] \\ &= \underline{\pi}^{-1}(t)[\hat{\underline{x}}(t|T) - \underline{\pi}(t)\underline{\lambda}(t|T)].\end{aligned} \tag{17b}$$

Observe that, via (14) and (7a)

$$\begin{aligned}&E[\hat{\underline{x}}_\beta(t|T)\underline{\lambda}'_\beta(t|T)] \\ &= \underline{P}_\beta(t|T)\underline{P}_\rho^{-1}(t|T)E(\hat{\underline{x}}(t|t) + \underline{\mathcal{O}}^{-1}(t|T)\lambda(t|T))(\hat{\underline{x}}(t|t) + (\underline{P}(t) - \underline{\pi}(t))\underline{\lambda}(t|T))'\underline{\pi}^{-1}(t) \\ &= \underline{P}_\beta(t|T)\underline{P}_\rho^{-1}(t|T)[\underline{\pi}(t) - \underline{P}(t) + \underline{\mathcal{O}}^{-1}(t|T)\underline{\mathcal{O}}(t|T)(\underline{P}(t) - \underline{\pi}(t))]\underline{\pi}^{-1}(t) \\ &= 0 .\end{aligned}$$

Since $E(\hat{\underline{x}}(t|t)\underline{\lambda}'(t|T)) = \underline{0}$; $E[\hat{\underline{x}}(t|t)\hat{\underline{x}}'(t|t)] = \underline{\pi}(t) - \underline{P}(t)$.

Consider then

$$E(\underline{x}(t) - \hat{\underline{x}}_\beta(t|T))\underline{y}'(s) = -\underline{P}_\rho(t|T)E[\underline{\lambda}_\beta(t|T)\underline{y}'(s)] \text{ since}$$

$$E(\underline{x}(t) - \hat{\underline{x}}(t|T))\underline{y}'(s) = \underline{0} \text{ for all observations } s \text{ in } [0,T] \tag{18}$$

$$= \underline{P}_\beta(t|T)E(\underline{\pi}^{-1}(t)\hat{\underline{x}}(t|T) - \underline{\lambda}(t|T))\underline{y}'(s)$$

$$= -\underline{P}_\beta(t|T)E[\underline{\pi}^{-1}(t)\underline{x}(t) - \underline{\lambda}(t|T)]\underline{y}'(s) \quad \text{by (18).}$$

Now substituting (1) in (18) gives

$$E[\underline{x}(t) - \hat{\underline{x}}(t|t) - \underline{P}(t)\underline{\lambda}(t|T)]\underline{y}'(s) = 0 \text{ for all } s \text{ in } [0,T]. \tag{19}$$

Thus

$$E(\underline{x}(t) - \hat{\underline{x}}_{\beta}(t|T))\underline{y}'(s) = -\underline{P}_{\beta}(t|T)E[\underline{\Pi}^{-1}(t)\underline{x}(t) - \underline{P}^{-1}(t)(\hat{\underline{x}}(t) - \hat{\underline{x}}(t|t))]\underline{y}'(s).$$

Now if, for s in $(t,T]$

$$E[\hat{\underline{x}}(t|t)\underline{y}'(s)] = E[\hat{\underline{x}}(t|t)\underline{x}'(t)]\underline{\Pi}^{-1}(t)\ E(\underline{x}(t)\underline{y}'(s)) \tag{20a}$$

then we find for s in $(t,T]$

$$\begin{aligned}
&E[\underline{x}(t)-\hat{\underline{x}}_{\beta}(t|T)]\underline{y}'(s) \\
&= -\underline{P}_{\beta}(t|T)[\underline{\Pi}^{-1}(t)\underline{\Pi}(t) - \underline{P}^{-1}(t)(\underline{\Pi}(t) - E[\hat{\underline{x}}(t|t)\underline{x}'(t)])\underline{\Pi}^{-1}(t)E(\underline{x}(t)\underline{y}'(s)) \\
&= \underline{P}_{\beta}(t|T)[\underline{I}-\underline{P}^{-1}(t)E(\underline{x}(t) - \hat{\underline{x}}(t|t))\underline{x}'(t)]\underline{\Pi}^{-1}(t)E(\underline{x}(t)\underline{y}'(s)) \\
&= -\underline{P}_{\beta}(t|T)[\underline{I} - \underline{P}^{-1}(t)\underline{P}(t)]\underline{\Pi}^{-1}(t)E(\underline{x}(t)\underline{y}'(s)) \\
&= \underline{0}\ .
\end{aligned}$$

Thus $\hat{\underline{x}}_{\beta}(t|T)$ is a the backwards least squares estimate of $\underline{x}(t)$ given the data in $(t,T]$ if (20a) holds.

It is now shown that (20a) holds if $\underline{x}(t),\underline{y}(t)$ are jointly "wide sense" Markov i.e. jointly Markov so far as second order statistics are concerned. To put this another way we show (20a) holds if

$$\hat{E}(\underline{y}(s)|\underline{\mathcal{Y}}_t,\underline{x}(t)) = \hat{E}(\underline{y}(s)|\underline{x}(t)) \qquad s \text{ in } (t,T] \tag{20b}$$

where, in discrete time $\underline{\mathcal{Y}}_t$ denotes the vector of data collected to time t ; while in continuous time $\underline{\mathcal{Y}}_t$ denotes the Hilbert space spanned by $\underline{y}(\tau)$ $0 \le \tau < t$. Also $\hat{E}$ denotes "wide sense" conditional expectation or projection (cf. Doob, [2, p150]; Parzen [12, p309]). The discrete and continuous time cases can now be treated jointly. With a slight abuse of notation denote (the Hilbert subspaces)

$$\hat{\underline{\mathcal{Y}}}_t = \hat{E}(\underline{\mathcal{Y}}_t|\underline{x}(t)); \quad \tilde{\underline{\mathcal{Y}}}_t = \underline{\mathcal{Y}}_t - \hat{\underline{\mathcal{Y}}}_t$$

e.g. $\hat{\mathcal{Y}}_t$ is the Hilbert space spanned by $\hat{E}(\underline{y}(s)|\underline{x}(t))$ s in $[0,t)$

(Recall that in discrete time $\hat{\underline{\mathcal{Y}}}_t = E(\underline{\mathcal{Y}}_t\underline{x}'(t))\underline{\Pi}^{-1}(t)\underline{x}(t))$.

Then consider that

$$\begin{aligned}
\hat{E}(\underline{y}(s)|\underline{\mathcal{Y}}_t,\underline{x}(t)) &= \hat{E}(\underline{y}(s)|\tilde{\underline{\mathcal{Y}}}_t,\underline{x}(t)) \\
&= \hat{E}(\underline{y}(s)|\tilde{\underline{\mathcal{Y}}}_t) + \hat{E}(\underline{y}(s)|\underline{x}(t))\ .
\end{aligned}$$

Thus (20b) holds if and only if

$$\hat{E}(\underline{y}(s)|\tilde{\mathcal{Y}}_t) = 0 \qquad T \geq s > t$$

This clearly holds if and only if

$$E(\underline{y}(s)\tilde{\underline{y}}'(\sigma)) = 0 \quad T \geq s > t \text{ and all observed points } \sigma \text{ in } [0,t)$$

i.e. if and only if

$$E(\underline{y}(s)\underline{y}'(\sigma)) = E(\underline{y}(s)\hat{\underline{y}}'(\sigma))$$

i.e. if and only if

$$E(\underline{y}(s)\underline{y}'(\sigma)) = E(\underline{y}(s)\underline{x}'(t))\underline{\pi}^{-1}(t)E(\underline{x}(t)\underline{y}'(\sigma))$$

$$\text{for } T \geq s > t \text{ and all observed } \sigma \text{ in } [0,t).$$

Now $\hat{\underline{x}}_t$ depends linearly on $\mathcal{Y}_t$ thus rewriting the last expression as

$$E(\mathcal{Y}_t\underline{y}'(s)) = E(\mathcal{Y}_t\underline{x}'(t))\underline{\pi}^{-1}(t)E(\underline{x}(t)\mathcal{Y}'(s))$$

we see that (20b) implies (20a). Perhaps the simplest way to visualize the joint wide sense Markovian requirement is in terms of a State Space Model for $\underline{x}(t)$, $\underline{y}(t)$. In the next section it is shown how, with a State Space Model, $\hat{\underline{x}}_\beta(t|T)$, $\hat{\underline{x}}_\rho(t|T)$ can be computed by backwards filters.

5. <u>BACKWARD FILTERS.</u> First continuous observations are discussed. Consider the state space model

$$d\underline{y}(t) = \underline{H}(t)\underline{x}(t)dt + \underline{v}(t)dt \qquad (21a)$$

$$\dot{\underline{x}}(t) = \underline{F}(t)\underline{x}(t) + \underline{G}(t)\underline{w}(t) \qquad (21b)$$

where $v(t)$, $w(t)$ are white noises satisfying

$$E\underline{v}(t)\underline{v}'(s) = \underline{\delta}(t-s), \qquad E(\underline{w}(t)\underline{w}'(s)) = \underline{\delta}(t-s)$$

$$E\underline{v}(t)\underline{w}'(s) = \underline{0} \qquad \forall\ s,t\ .$$

According to the orthogonality condition the backwards filtered estimate is given by (the subscript β is now replaced by b)

$$\hat{\underline{x}}_b(t|T) = \int_t^T E(\underline{x}(t)\underline{\nu}_b'(s|T))\underline{\nu}_b(s|T)ds \qquad (22a)$$

where $\underline{\nu}_b(s|T)$ is the backwards innovations (i.e. over $[0,T]$ $\underline{\nu}_b(s|T)ds$ is linearly invertibly equivalent to the data $d\underline{y}(\sigma)$. i.e. the Hilbert space spanned by "$\underline{\nu}_b(s|T)ds$" is the same as the one spanned by "$d\underline{y}(\sigma)$").

We now find the backwards Kalman Filter. First

$$\frac{d\hat{\underline{x}}_b(t|T)}{dt} = -E(\underline{x}(t)\underline{\nu}_b'(t|T))\underline{\nu}_b(t|T) + \int_t^T d/dt\, E(\underline{x}(t)\underline{\nu}_b'(s|T))\underline{\nu}_b(s|T)ds. \qquad (22b)$$

To compute terms such as

$$\underline{\nu}_b(t|T),\ E[\underline{x}(t)\underline{\nu}_b'(t|T)],\ d/dt\,E[((\underline{x}(t)\underline{\nu}_b'(s|T))]$$

we need to <u>reorganize</u> equations (21a) and (21b) into a backwards model where the noises are orthogonal to future values of $\underline{x}(t)$. Recently Verghese and Kailath [13] have shown how this can be done.

In Appendix B it is shown that the following filter results

$$d\hat{\underline{x}}_b(t|T) = \underline{F}_b(t)\hat{\underline{x}}_b(t|T)dt - \underline{P}_b(t|T)\underline{H}'(t)\underline{\nu}_b(t|T)dt \qquad (23a)$$

$$\underline{\nu}_b(t|T)dt = d\underline{y}(t) - \underline{H}(t)\hat{\underline{x}}_b(t|T)dt \qquad (23b)$$

with initial condition $\hat{\underline{x}}_b(T|T) = \underline{0}$ where

$$\underline{F}_b(t) = \underline{F}(t) + \underline{G}(t)\underline{G}'(t)\underline{\pi}^{-1}(t) = -\underline{\pi}(t)\underline{F}'(t)\underline{\pi}^{-1}(t). \qquad (23c)$$

Also

$$-d\underline{P}_b(t|T)/dt = -\underline{P}_b(t|T)\underline{F}_b'(t) - \underline{F}_b(t)\underline{P}_b(t|T) + G(t)\underline{G}'(t) - \underline{P}_b(t|T)\underline{H}'(t)\underline{H}(t)\underline{P}_b(t|T) \quad (24a)$$

with initial condition $\underline{P}_b(T|T) = \underline{\pi}(T)$ alternatively

$$d\underline{P}_b^{-1}(t|T)/dt = -\underline{P}_b^{-1}(t|T)\underline{F}_b(t) - \underline{F}'_b(t)\underline{P}_b^{-1}(t|T) - \underline{H}'(t)\underline{H}(t)$$
$$+ \underline{P}_b^{-1}(t|T)\underline{G}(t)\underline{G}'(t)\underline{P}_b^{-1}(t|T). \tag{24b}$$

Now equations (23), (24) are not the backwards filter equations that are usually given. The equations are usually given for

$$\underline{z}(t|T) = \underline{P}_b^{-1}(t|T)\hat{\underline{x}}_b(t|T) \tag{25a}$$

or for

$$\hat{\underline{x}}_r(t|T) = \underline{P}_r(t|T)\underline{z}(t|T). \tag{25b}$$

In Appendix C it is shown that filters for these quantities are

$$d\underline{z}(t|T) = -(\underline{F}'(t) - \underline{P}_r^{-1}(t|T)\underline{G}(t)\underline{G}'(t))\underline{z}(t|T)dt - \underline{H}'(t)d\underline{y}(t) \tag{26a}$$

$$d\hat{\underline{x}}_r(t|T) = (\underline{F}(t) + \underline{P}_r(t|T)\underline{H}'(t)\underline{H}(t))\hat{\underline{x}}_r(t|T)dt - \underline{P}_r(t|T)\underline{H}'(t)d\underline{y}(t) \tag{26b}$$

$$= \underline{F}(t)\hat{\underline{x}}_r(t|T)dt + \underline{P}_r(t|T)\underline{H}'(t)\underline{\nu}_r(t|T)dt \tag{26c}$$

with $\underline{\nu}_r(t|T)dt = d\underline{y}(t) - \underline{H}(t)\hat{\underline{x}}_r(t|T)dt$ and initial condition arbitrary. Also

$$d\underline{P}_r^{-1}(t|T)/dt = -\underline{P}_r^{-1}(t|T)\underline{F}(t) - \underline{F}'(t)\underline{P}_r^{-1}(t|T) - \underline{H}'(t)\underline{H}(t)$$
$$+ \underline{P}_r^{-1}(t|T)\underline{G}(t)\underline{G}'(t)\underline{P}_r^{-1}(t|T) \tag{26d}$$

with initial condition $\underline{P}_r^{-1}(T|T) = \underline{0}$.

Equations (26a), (26b), (26d) appear for example in Ljung and Kailath [9, respectively equations (14), (16), (13)] ("b" in their notation is equivalent to "r" of the present notation).

6. <u>SOME ADDITIONAL TWO-FILTER-LIKE REPRESENTATIONS.</u> In Section 4 it was pointed out how there is a pseudo-backwards expression analogous to (1) namely

$$\underline{x}(t|T) = \underline{x}_\beta(t|T) + \underline{P}_\beta(t|T)\underline{\lambda}_\beta(t|T) \tag{30}$$

where also

$$E(\hat{\underline{x}}_\beta(t|T)\underline{\lambda}'_\beta(t|T)) = \underline{0}\ .$$

If we define

$$\underline{\mathcal{O}}_\beta(t|T) = E(\underline{\lambda}_\beta(t|T)\underline{\lambda}'_\beta(t|T)) \tag{32}$$

then it follows from (30) that

$$\begin{aligned}\underline{P}_\beta(t|T) &= E(\underline{x}(t) - \hat{\underline{x}}_\beta(t|T))(\underline{x}(t) - \hat{\underline{x}}_\beta(t|T))' \\ &= \underline{P}(t|T) + \underline{P}_\beta(t|T)\underline{\mathcal{O}}_\beta(t|T)\underline{P}_\beta(t|T).\end{aligned} \tag{33}$$

Thus we have pseudo-backwards analogues of (1)-(4).

Now however we can retrace the argument of Section 3 to produce an estimate

$$\hat{\underline{x}}_\phi(t|T) = \hat{\underline{x}}_\beta(t|T) + \underline{\mathcal{O}}_\beta^{-1}(t|T)\underline{\lambda}_\beta(t|T) \tag{34}$$

(where the subscript ϕ denotes forward) satisfying

$$E(\underline{x}(t) - \hat{\underline{x}}_\beta(t|T))(\underline{x}'(t) - \hat{\underline{x}}'_\phi(t|T)) = \underline{0} \tag{35a}$$

$$E(\underline{x}(t) - \hat{\underline{x}}_\beta(t|T))\underline{x}'(t) = \underline{0} \tag{35b}$$

with also

$$\underline{\mathcal{O}}_\beta^{-1}(t|T) = \underline{P}_\beta(t|T) + \underline{P}_\phi(t|T) \tag{36}$$

where

$$\underline{P}_\phi(t|T) = E(\underline{x}(t) - \hat{\underline{x}}_\phi(t|T))(\underline{x}(t) - \hat{\underline{x}}_\phi(t|T))'.$$

Further

$$\underline{P}^{-1}(t|T) = \underline{P}_\beta^{-1}(t|T) + \underline{P}_\phi^{-1}(t|T). \tag{37a}$$

Now substituting (13) in (37a) while equating (37a) to (6) gives

$$\underline{P}^{-1}(t) = \underline{\pi}^{-1}(t) + \underline{P}_\phi^{-1}(t|T). \tag{37b}$$

Then we could search for $\hat{\underline{x}}_{\beta\beta}(t|T)$ satisfying

$$E(\underline{x}(t) - \hat{\underline{x}}_{\beta\beta}(t|T))\hat{\underline{x}}'_{\beta\beta}(t|T) = \underline{0} \tag{38}$$

and find

$$\hat{\underline{x}}_{\beta\beta}(t|T) = \underline{\pi}(t)(\underline{\pi}(t) + \underline{P}_\phi(t|T))^{-1}\hat{\underline{x}}_\phi(t|T) \tag{39}$$

as well as

$$\underline{P}_{\phi}^{-1}(t|T)\hat{\underline{x}}_{\phi}(t|T) = \underline{P}_{\beta\beta}^{-1}(t|T)\hat{\underline{x}}_{\beta\beta}(t|T) \tag{40}$$

where

$$\underline{P}_{\beta\beta}(t|T) = E(\underline{x}(t) - \hat{\underline{x}}_{\beta\beta}(t|T))\underline{x}'(t).$$

However we find in Appendix D that

$$\hat{\underline{x}}_{\beta\beta}(t|T) = \hat{\underline{x}}(t|t) \tag{41a}$$

so

$$\underline{P}_{\beta\beta}(t|T) = \underline{P}(t). \tag{41b}$$

If $\hat{\underline{x}}_b(t|T)$ denotes the linear least squares estimate of $\underline{x}(t)$ given the data in $[t,T]$ then again all the equations just derived (except (37b), (41)) hold with β replaced by b and ϕ replaced by f (in (37b) replace $\underline{P}^{-1}(t)$ by $\underline{P}_{bb}^{-1}(t|T)$). Then we can also discover the wide sense Markov condition of Section 5 by requiring that

$$\hat{\underline{x}}_{bb}(t|T) = \underline{x}(t|t) .$$

7. Conclusion. By simple argument, a number of two-filter-like formulae, that apply to fairly general nonstationary processes, have been derived for the smoothing problem of linear estimation. In general the filters cannot be interpreted in a backwards sense unless the joint signal/observations model is wide sense Markovian.

Appendix A. Derivation of (17).

From (14)

$$\hat{\underline{x}}_\beta(t|T) = \underline{P}_\beta(t|T)\underline{P}_\rho^{-1}(t|T)\hat{\underline{x}}_\rho(t|T)$$

$$= \underline{P}_\beta(t|T)\underline{P}_\rho^{-1}(t|T)(\hat{\underline{x}}(t|t) + \underline{\mathcal{O}}^{-1}(t|T)\underline{\lambda}(t|T)) \qquad \text{by (7a)}$$

$$= \underline{P}_\beta(t|T)(\underline{P}_\rho^{-1}(t|T) + \underline{\pi}^{-1}(t))\hat{\underline{x}}(t|t) - \underline{P}_\beta(t|T)\underline{\pi}^{-1}(t)\hat{\underline{x}}(t|t)$$
$$+ \underline{P}_\beta(t|T)\underline{P}_\rho^{-1}(t|T)\underline{\mathcal{O}}^{-1}(t|T)\underline{\lambda}(t|T)$$

$$= \hat{\underline{x}}(t|t) + \underline{P}_\beta(t|T)[\underline{P}_\rho^{-1}(t|T)\underline{\mathcal{O}}^{-1}(t|T)\underline{\lambda}(t|T) - \underline{\pi}^{-1}(t)\hat{\underline{x}}(t|t)] \qquad \text{by (13)}$$

$$= \hat{\underline{x}}(t|T) - \underline{P}_\beta(t|T)[\underline{\pi}^{-1}(t)\hat{\underline{x}}(t|t)$$
$$- \underline{P}_\rho^{-1}(t|T)\underline{\mathcal{O}}^{-1}(t|T)\underline{\lambda}(t|T) + \underline{P}_\beta^{-1}(t|T)\underline{P}(t)\underline{\lambda}(t|T)]$$

$$= \hat{\underline{x}}(t|T) - \underline{P}_\beta(t|T)[\underline{\pi}^{-1}(t)\hat{\underline{x}}(t|t) + \underline{\pi}^{-1}(t)(\underline{P}(t) - \underline{\pi}(t))\underline{\lambda}(t|T)] .$$

Since

$$\underline{P}_\beta^{-1}(t|T)\underline{P}(t) - \underline{P}_\rho^{-1}(t|T)\underline{\mathcal{O}}^{-1}(t|T)$$

$$= (\underline{P}_\rho^{-1}(t|T) + \underline{\pi}^{-1}(t))\underline{P}(t) - \underline{P}_\rho^{-1}(t|T)\underline{\mathcal{O}}^{-1}(t|T) \qquad \text{by (13)}$$

$$= \underline{P}_\rho^{-1}(t|T)(\underline{P}(t) - \underline{\mathcal{O}}^{-1}(t|T)) + \underline{\pi}^{-1}(t)\underline{P}(t)$$

$$= -\underline{I} + \underline{\pi}^{-1}(t)\underline{P}(t) \qquad \text{by (5).}$$

Thus

$$\hat{\underline{x}}(t|T) = \hat{\underline{x}}_\beta(t|T) + \underline{P}_\beta(t|T)\underline{\pi}^{-1}(t)[\hat{\underline{x}}(t|t) + (\underline{P}(t) - \underline{\pi}(t))\underline{\lambda}(t|T)].$$

Appendix B. Derivation of (23a), (23b), (24a), (24b).

In the present situation the backwards model is

$$d\underline{y}(t) = \underline{H}(t)\underline{x}(t)dt + \underline{\tilde{v}}(t)dt$$

$$\dot{\underline{x}}(t) = \underline{x}(t) + \underline{G}(t)\underline{\tilde{w}}(t)$$

where

$$E(\underline{\tilde{v}}(t)\underline{\tilde{v}}(s)) = \delta(t-s) \ ; \quad E(\underline{\tilde{w}}(t)\underline{\tilde{w}}'(s)) = \delta(t-s)$$

$$E(\underline{\tilde{v}}(t)\underline{\tilde{w}}(s)) = \underline{0} \quad \forall\ t,s$$

$$E(\underline{\tilde{v}}(t)\underline{y}'(s)) = \underline{0} \ ; \quad E(\underline{\tilde{w}}(t)\underline{y}'(s)) = \underline{0} \qquad s > t\ .$$

Also

$$\underline{F}_b(t) = \underline{F}(t) + \underline{G}(t)\underline{G}'(t)\underline{\pi}^{-1}(t) = -\underline{\pi}(t)\underline{F}'(t)\underline{\pi}^{-1}(t) \tag{C1}$$

so that

$$\underline{F}_b(t)\underline{\pi}(t) = -\underline{\pi}(t)\underline{F}'(t). \tag{C2}$$

It follows that

$$\underline{\nu}_b(t|T)dt = d\underline{y}(t) - \hat{\underline{y}}_b(t|T)dt = d\underline{y}(t) - \underline{H}(t)\hat{\underline{x}}_b(t|T)dt \tag{23b}$$

$$E[\underline{x}(t)\underline{\nu}_b'(t|T)] = E\{\underline{x}(t)[\underline{H}(t)(\underline{x}(t) - \hat{\underline{x}}_b(t|T)) + \underline{\tilde{v}}(t)]\}'$$

$$= \underline{P}_b(t|T)\underline{H}'(t) \tag{C4}$$

$$d/dt\ E[\underline{x}(t)\underline{\nu}_b'(s|T)] = \underline{F}_b(t)E(\underline{x}(t)\underline{\nu}_b'(s|T)) \tag{C5}$$

Thus in (22b)

$$d\hat{\underline{x}}_b(t|T) = \underline{F}_b(t)\hat{\underline{x}}_b(t|T)dt - \underline{P}_b(t|T)H'(t)(d\underline{y}(t) - \underline{H}(t)\hat{\underline{x}}_b(t|T))dt \tag{C6}$$

$$= (\underline{F}_b(t) + \underline{P}_b(t|T)\underline{H}'(t)\underline{H}(t))\hat{\underline{x}}_b(t|T)dt - \underline{P}_b(t|T)\underline{H}'(t)d\underline{y}(t). \tag{C7}$$

Equation (23a) follows from (C6) and (23b). Next (22a) implies

$$\underline{\pi}(t) - \underline{P}_b(t|T) = E(\hat{\underline{x}}_b(t|T)\hat{\underline{x}}_b'(t|T))$$

$$= \int_t^T E[\underline{x}(t)\underline{\nu}'_b(s|T)]E(\underline{\nu}_b(s|T)\underline{x}'(t))ds \tag{C8}$$

Notice that $\underline{P}_b(T|T) = \underline{\pi}(T)$.

Thus differentiating and using (C4) gives (using "$\cdot$" for "d/dt")

$$\dot{\underline{\pi}} - \dot{\underline{P}}_b(t|T) = -\underline{P}_b(t|T)\underline{H}'(t)\underline{H}(t)\underline{P}_b(t|T) + \underline{F}_b(t)(\underline{\pi}(t) - \underline{P}_b(t|T))$$

$$+ (\underline{\pi}(t) - \underline{P}_b(t|T))\underline{F}_b'(t).$$

Of course

$$\dot{\underline{\pi}} = \underline{F}(t)\underline{\pi}(t) + \underline{\pi}(t)\underline{F}'(t) + \underline{G}(t)\underline{G}'(t).$$

Thus using (C2) gives

$$-\dot{\underline{P}}_b(t|T) = -\underline{P}_b(t|T)\underline{H}'(t)\underline{H}(t)\underline{P}_b(t|T) - \underline{F}_b(t)\underline{P}_b(t|T) - \underline{P}_b(t|T)\underline{F}'_b(t) + \underline{G}(t)\underline{G}'(t). \tag{24a}$$

Also it follows that

$$\begin{aligned} d/dt\ \underline{P}_b^{-1}(t|T) &= -\underline{P}_b^{-1}(t|T)\dot{\underline{P}}_b(t|T)\underline{P}_b^{-1}(t|T) \\ &= -\underline{P}_b^{-1}(t|T)\underline{F}_b(t) - \underline{F}'_b(t)\underline{P}_b^{-1}(t|T) - \underline{H}(t)'\underline{H}(t) \\ &\quad + \underline{P}_b^{-1}(t|T)\underline{G}(t)\underline{G}'(t)\underline{P}_b^{-1}(t|T) \end{aligned} \tag{24b}$$

Appendix C. Derivation of (26a), (26b), (26d).

Differentiate $\underline{z}(t|T) = \underline{P}_b^{-1}(t|T)\hat{\underline{x}}_b(t|T)$ to find

$$d\underline{z}(t|T)/dt = \underline{P}_b^{-1}(t|T)\,d\hat{\underline{x}}_b(t|T)/dt + d\underline{P}_b^{-1}(t|T)/dt\,\hat{\underline{x}}_b(t|T)$$

$$= \underline{P}_b^{-1}(t|T)(\underline{F}_b(t) + \underline{P}_b(t|T)\underline{H}'(t)\underline{H}(t))\hat{\underline{x}}_b(t|T) - \underline{H}'(t)d\underline{y}(t)/dt$$

$$-(\underline{P}_b^{-1}(t|T)\underline{F}_b(t) - \underline{F}_b'(t)\underline{P}_b^{-1}(t|T) + \underline{H}'(t)\underline{H}(t)$$

$$-\underline{P}_b^{-1}(t|T)\underline{G}(t)\underline{G}'(t)\underline{P}_b^{-1}(t|T))\hat{\underline{x}}_b(t|T)$$

by (23a), (23b), (24a)

$$= -(\underline{F}_b'(t) - \underline{P}_b^{-1}(t)\underline{G}(t)\underline{G}'(t))\underline{P}_b^{-1}(t|T)\hat{\underline{x}}_b(t|T) - \underline{H}'(t)d\underline{y}(t)/dt$$

$$= -(\underline{F}_b'(t)\ \underline{\pi}^{-1}(t)\underline{G}(t)\underline{G}'(t) - \underline{P}_r^{-1}(t|T)\underline{G}(t)\underline{G}'(t))\underline{z}(t|T) - \underline{H}'(t)d\underline{y}(t)/dt$$

by (13)

$$= -(\underline{F}'(t) - \underline{P}_r^{-1}(t|T)\underline{G}(t)\underline{G}'(t))\underline{z}(t|T) - \underline{H}'(t)d\underline{y}(t)/dt$$

by (C2).

This is (26a). For (26d) begin with (13)

$$\underline{P}_r^{-1}(t|T) = \underline{P}_b^{-1}(t|T) - \underline{\pi}^{-1}(t) .$$

Notice that

$$d/dt\ \underline{\pi}^{-1}(t) = -(\underline{\pi}^{-1}(t)\underline{F}(t) + \underline{F}'(t)\underline{\pi}^{-1}(t) + \underline{\pi}^{-1}(t)\underline{G}(t)\underline{G}'(t)\underline{\pi}^{-1}(t)).$$

Thus

$$\begin{aligned} d/dt\ \underline{P}_r^{-1}(t|T) = {} & -\underline{P}_b^{-1}(t|T)\underline{F}_b(t) - \underline{F}_b'(t)\underline{P}_b^{-1}(t|T) - \underline{H}'(t)\underline{H}(t) \\ & + \underline{P}_b^{-1}(t|T)\underline{G}(t)\underline{G}'(t)\underline{P}_b^{-1}(t|T) \\ & - \underline{\pi}^{-1}(t)\underline{F}(t) - \underline{F}'(t)\underline{\pi}^{-1}(t) - \underline{\pi}^{-1}(t)\underline{G}(t)\underline{G}'(t)\underline{\pi}^{-1}(t) \\ = {} & -\underline{P}_b^{-1}(t|T)\underline{F}_b(t) - \underline{F}_b'(t)\underline{P}_b^{-1}(t|T) - \underline{H}'(t)\underline{H}(t) \\ & + \underline{P}_b^{-1}(t|T)\underline{G}(t)\underline{G}'(t)\underline{P}_b^{-1}(t|T) + \underline{F}_b'(t)\underline{\pi}^{-1}(t) + \underline{\pi}^{-1}(t)\underline{F}_b(t) \\ & - \underline{\pi}^{-1}(t)\underline{G}(t)\underline{G}'(t)\underline{\pi}^{-1}(t) \end{aligned}$$

by (C2).

Now apply (13) to find

$$= -\underline{P}_r^{-1}(t|T)\underline{F}_b(t) - \underline{F}_b'(t)\underline{P}_r^{-1}(t|T) - \underline{H}'(t)\underline{H}(t)$$

$$+ \underline{P}_b^{-1}(t|T)\underline{G}(t)\underline{G}'(t)\underline{P}_b^{-1}(t|T)$$

$$= -\underline{P}_r^{-1}(t|T)(\underline{F}(t) + \underline{G}(t)\underline{G}'(t)\underline{\pi}^{-1}(t))$$

$$-(\underline{F}'(t) + \underline{\pi}^{-1}(t)\underline{G}(t)\underline{G}'(t))\underline{P}_r^{-1}(t|T) - \underline{H}'(t)\underline{H}(t)$$

$$+ \underline{P}_r^{-1}(t|T)\underline{G}(t)\underline{G}'(t)\underline{P}_r^{-1}(t|T) + \underline{\pi}^{-1}(t)\underline{G}(t)\underline{G}'(t)\underline{P}_r^{-1}(t|T)$$

$$+ \underline{P}_r^{-1}(t|T)\underline{G}(t)\underline{G}'(t)\underline{\pi}^{-1}(t) + \underline{\pi}^{-1}(t)\underline{G}(t)\underline{G}'(t)\underline{\pi}^{-1}(t)$$

$$= -\underline{P}_r^{-1}(t|T)\underline{F}(t) - \underline{F}'(t)\underline{P}_r^{-1}(t|T) + \underline{P}_r^{-1}(t|T)\underline{G}(t)\underline{G}'(t)\underline{P}_r^{-1}(t|T) - \underline{H}'(t)\underline{H}(t)$$

which is (26d)

Clearly also

$$d/dt\underline{P}_r(t|T) = -\underline{P}_r(t|T)d\underline{P}_r^{-1}(t|T)/dt\ \underline{P}_r(t|T)$$

$$= \underline{F}(t)\underline{P}_r(t|T) + \underline{P}_r(t|T)\underline{F}'(t)\underline{G}(t)\underline{G}'(t)$$

$$+ \underline{P}_r(t|T)\underline{H}'(t)\underline{H}(t)\underline{P}_r(t|T).$$

From this expression, (26d) and differentiating in (25b), then (26b) is easily established.

Appnedix D. Derivation of (41a)

Consider

$$
\begin{aligned}
\hat{\underline{x}}_{\beta\beta}(t|T) &= \underline{\pi}(t)(\underline{\pi}(t) + \underline{P}_{\phi}(t|T))^{-1}\hat{\underline{x}}_{\phi}(t|T) \\
&= \underline{\pi}(t)(\underline{\pi}(t) + \underline{P}_{\phi}(t|T))^{-1}(\hat{\underline{x}}_{\beta}(t|T) + \underline{\mathcal{O}}_{\beta}^{-1}\underline{\lambda}_{\beta}(t|T)) \\
&\qquad\qquad \text{by (34)} \\
&= \underline{\pi}(t)(\underline{\pi}(t) + \underline{P}_{\phi}(t|T))^{-1}(\hat{\underline{x}}(t|T) - (\underline{P}_{\beta}(t|T) - \underline{\mathcal{O}}_{\beta}^{-1}(t|T))\underline{\lambda}_{\beta}(t|T)) \\
&\qquad\qquad \text{by (30)} \\
&= \underline{\pi}(t)(\underline{\pi}(t) + \underline{P}_{\phi}(t|T))^{-1}(\hat{\underline{x}}(t|T) + \underline{P}_{\phi}(t|T)\underline{\lambda}_{\beta}(t|T)) \\
&\qquad\qquad \text{by (36)} \\
&= (\underline{\pi}^{-1}(t) + \underline{P}_{\phi}^{-1}(t|T))^{-1}(\underline{P}_{\phi}^{-1}(t|T)\hat{\underline{x}}(t|T) + \underline{\lambda}_{\beta}(t|T)) \\
&= P(t)[(\underline{P}^{-1}(t) - \underline{\pi}^{-1}(t))\hat{\underline{x}}(t|T) + \underline{\lambda}_{\beta}(t|T)] \\
&\qquad\qquad \text{by (37b) twice} \\
&= \underline{P}(t)[(\underline{P}^{-1}(t) - \underline{\pi}^{-1}(t))\hat{\underline{x}}(t|T) + \underline{\pi}^{-1}(t)\hat{\underline{x}}(t|T) - \underline{\lambda}(t|T)] \\
&\qquad\qquad \text{by (17b)} \\
&= \underline{P}(t)[\underline{P}^{-1}(t)\hat{\underline{x}}(t|T) - \underline{\lambda}(t|T)] \\
&= \underline{P}(t)[\underline{P}^{-1}(t)(\hat{\underline{x}}(t|t) + \underline{P}(t)\underline{\lambda}(t|T)) - \underline{\lambda}(t|T)] \\
&\qquad\qquad \text{by (1)} \\
&= \hat{\underline{x}}(t|t).
\end{aligned}
$$

REFERENCES

1. F. A. Badawi, A. Lindquist and M. Pavon, "A Stochastic Realization Approach to the Smoothing Problem", IEEE Trans. Autom. Contr., Vol. AC-24, pp. 878-888, 1979.

2. J. L. Doob, "Stochastic Processes", New York, John Wiley, 1953.

3. D. G. Fraser, A New Technique for Optimal Smoothing of Data, Sc. D. Dissertation, M.I.T., 1967.

4. T. Kailath, "Likelihood Ratios for Gaussian Processes", IEEE Trans. Inf. Thy., Vol. IT-16, pp. 276-287, 1970.

5. T. Kailath, "A Note on Least Squares Estimation by the Innovations Method", SIAM J. Contr., Vol. 10, pp. 477-486, 1972.

6. T. Kailath, P. Frost, "An Innovation Approach to Least-Squares Estimation, Part II: Linear Smoothing in Additive White Noise, IEEE Trans. Autom. Contr., Vol. AC-13, pp. 655-660, 1968.

7. T. Kailath and R. A. Geesey, "An Innovations Approach to Least-Squares Estimation, Part V'. Innovations Representations and Recursive Estimation in Coloured Noise", IEEE Trans. Autom. Contr., Vol. AC-18, pp. 435-453, 1973.

8. D. G. Lainiotis, "A Unifying Framework for Linear Estimation: Generalized Partitioned Algorithms", Inf. Sci., Vol. 10, pp. 243-278, 1976.

9. L. Ljung and T. Kailath, "A Unified Approach to Smoothing Formulas" Automatica, Vol. 12, pp. 147-157, 1976.

10. D. Q. Mayne, "A Solution of the Smoothing Problem for Linear Dynamic Systems", Automatica, Vol. 4, pp. 73-92, 1966.

11. R. K. Mehra, "On Optimal and Suboptimal Linear Smoothing", Tech. Rep. No. 559, Div. Eng. Appl. Phys, Harvard Univ., 1968.

12. E. Parzen, "Statistical Inference On Time Series by Hilbert Space Methods, I., Tech. Rep. Now. 23, Dept. Stats., Stanford Univ., 1959, Reprinted in "Time Series Analysis Papers" by E. Parzen, Holden-Day, 1967.

13. G. Verghese and T. Kailath, "A Further Note on Backwards Markovian Models", IEEE Trans. Inf. Thy., Vol. IT-25, pp. 121-124, 1979.

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER 2120

2. GOVT ACCESSION NO. AD-A093634

3. RECIPIENT'S CATALOG NUMBER (9) Technical

4. TITLE (and Subtitle)
(6) SMOOTHING ESTIMATION OF STOCHASTIC PROCESSES. PART II. TWO FILTER FORMULAE.

5. TYPE OF REPORT & PERIOD COVERED Summary Report - no specific reporting period

6. PERFORMING ORG. REPORT NUMBER

7. AUTHOR(s)
(10) V. Solo (12) 25

8. CONTRACT OR GRANT NUMBER(s)
(15) DAAG29-80-C-0041

9. PERFORMING ORGANIZATION NAME AND ADDRESS
Mathematics Research Center, University of Wisconsin
610 Walnut Street
Madison, Wisconsin 53706

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
Work Unit Number 4 - Statistics and Probability

11. CONTROLLING OFFICE NAME AND ADDRESS
U. S. Army Research Office
P.O. Box 12211
Research Triangle Park, North Carolina 27709

12. REPORT DATE (11) Sept[redacted]80

13. NUMBER OF PAGES 21

14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)
(14) MRC-TSR-2120

15. SECURITY CLASS. (of this report) UNCLASSIFIED

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)
Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)
Linear estimation, smoothing, filtering, Markov process, covariance kernel, innovations

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Some simple derivations of two-filter-like formulae (in the smoothing problem of linear estimation) are given for general nonstationary process. It then becomes clear how a wide sense Markovian assumption is required to give the formulae a backwards filter interpretation.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

221200

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

END

DATE
FILMED

2 - 81

DTIC